Canadian Technical Report of
Fisheries and Aquatic Services
No. 1531

February 1987

Partitioning Size from Morphometric Data: A Comparison of Five Statistical Procedures used in Fisheries Stock Identification Research

by

Ross R. Claytor
Science Branch, Gulf Region
Department of Fisheries and Oceans
P.O. Box 5030, Moncton, NB E1C 9B6

Hugh R. MacCrimmon
Department of Zoology
University of Guelph
Guelph, ON N1G 2W1

Fisheries and Oceans Pêches et Océans

Canada

87-04759/

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of the Department of Fisheries and Oceans, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in *Aquatic Sciences and Fisheries Abstracts* and indexed in the Department's annual index to scientific and technical publications.

Numbers 1–456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457–714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715–924 were issued as Department of Fisheries and the Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out-of-stock reports will be supplied for a fee by commercial agents.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. Il n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques du ministère des Pêches et des Océans, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications complètes. Le titre exact paraît au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la revue *Résumés des sciences aquatiques et halieutiques*, et ils sont classés dans l'index annual des publications scientifiques et techniques du Ministère.

Les numéros 1 à 456 de cette série ont été publiés à titre de rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

Canadian Technical Report of
Fisheries and Aquatic Sciences No. 1531

February 1987

PARTITIONING SIZE FROM MORPHOMETRIC DATA: A COMPARISON OF FIVE STATISTICAL PROCEDURES USED IN FISHERIES STOCK IDENTIFICATION RESEARCH

by

Ross R. Claytor
Science Branch, Gulf Region
Department of Fisheries and Oceans
P.O. Box 5030, Moncton, NB E1C 9B6

Hugh R. MacCrimmon
Department of Zoology
University of Guelph
Guelph, ON N1G 2W1

(c) Minister of Supply and Services Canada 1987

Cat. No. Fs 97-6/1531E ISSN 0706-6457

Correct citation for this publication is:

Claytor, R.R. and H.R. MacCrimmon. 1986. Partitioning size from morphometric data: a comparison of five statistical procedures used in fisheries stock identification research. Can. Tech. Rep. Fish. Aquat. Sci. No. 1531: viii + 23 p.

CONTENTS

List of Tables iv

List of Figures vi

Abstract/Résumé vii

Introduction 1

Materials and Methods 1

Results 2

Discussion 3

Conclusions 4

Acknowledgements 4

References 4

LIST OF TABLES

Table 1. Acronyms used for morphometric characters, geographic stocks, data sets, analyses, and symbols in subsequent tables and figures. 6

Table 2. R-squared values for regressions of size-free characters against size standard associated with each procedure. * indicates a non-significant regression (P > 0.001). Variable refers to either canonical variable or principal component. Acronyms as listed in Table 1. 7

Table 3. Kappa and average observed percentage of specimens correctly identified to size groups and geographic stocks by discriminant function analysis. Values for complete data set are for geographic stocks. Acronyms as listed in Table 1. 7

Table 4. The standardized coefficients of the first two canonical variables from the discriminant analysis using characters adjusted for size by total and pooled within-group regression slopes. Blank indicates that the character did not enter the model. Acronyms as listed in Table 1. A) First canonical variable; B) Second canonical variable. 8

Table 5. Partial correlation coefficients between morphometric characters before and after adjusting for size using total regression slopes for the non-overlapping, partial overlapping and complete overlapping data sets. Upper triangular matrix after size adjustment, lower triangular prior to size adjustment. Acronyms as listed in Table 1. A) Non-overlap; B) Partial overlap; C) Complete overlap. 9

Table 6. First principal component coefficients derived using the total covariance matrix and the within-group covariance matrix from all three data sets. Acronyms as listed in Table 1. 11

Table 7. The second and third principal component coefficients derived using the total covariance matrix, within-group covariance matrix and shear procedure for A) Non-overlapping data set; B) Partial overlapping data set; and C) Complete overlapping data set. Acronyms as listed in Table 1. 12

Table 8. Coefficients based on characters identified as most differentiated between stocks by principal component analysis of the entire character set. A) The four most differentiated characters identified by principal component analysis using the total covariance matrix. B) The three most differentiated characters identified by principal component analysis using the within-group covariance matrix. Acronyms as listed in Table 1. 14

LIST OF FIGURES

Fig. 1. The organization of the Canadian and Scottish stocks into data sets composed of non-overlapping, partially and completely overlapping size subgroups. 15

Fig. 2 Discriminant analysis using characters adjusted for size by the total overall regression slopes for all three data sets. Acronyms as listed in Table 1. 16

Fig. 3 Discriminant analysis using characters adjusted for size by the pooled within-group slopes for all three data sets. Acronyms as listed in Table 1. 17

Fig. 4 Principal component analysis using the total covariance matrix for non-overlapping and completely overlapping data sets. Acronyms as listed in Table 1. 18

Fig. 5 Principal component analysis using the total covariance matrix for partial overlapping data set. Acronyms as listed in Table 1. 19

Fig. 6 Principal component analysis using the pooled within-group covariance matrix for all three data sets. Acronyms as listed in Table 1. 20

Fig. 7 Principal component analysis using the shear procedure for non-overlapping data set. Acronyms as listed in Table 1. 21

Fig. 8 Principal component analysis of the completely overlapping data set using the four most differentiated characters identified by principal component analysis using the total covariance matrix. Acronyms as listed in Table 1. 22

Fig. 9 Principal component analysis of the completely overlapping data set using the three most differentiated characters identified by principal component analysis using the within covariance matrix. Acronyms as listed in Table 1. 23

ABSTRACT

Claytor, R.R. and H.R. MacCrimmon. 1986. Partitioning size from morphometric data: a comparison of five statistical procedures used in fisheries stock identification research. Can. Tech. Rep. Fish. Aquat. Sci. No. 1531: viii + 23 p.

Five statistical procedures currently used to partition size from morphometric data were compared using field data sets of non-overlapping, partially overlapping, and completely overlapping size groups. Procedures utilizing total rather than within-group variation were most consistent in partitioning size, identifying stocks, and character selection. These procedures were (i) discriminant function analysis using characters adjusted by the total overall slopes and (ii) principal component analysis using the total covariance matrix as input. The other procedures examined, (iii) discriminant function analysis using characters adjusted by the pooled within-group slopes and (iv) principal component analysis using either the within-group covariance matrix or the (v) shear procedure did not partition size from the non-overlapping data set. We conclude that total variation procedures are recommended when the proposed stocks consist of subgroups varying in size. However, if there is at least some size overlap among the stocks examined, then investigations utilizing discriminant function analysis with either total or pooled within-group slopes to adjust characters or investigations using principal component analysis based on the total covariance matrix or shear procedure are very similar. Adjustments for size should be made prior to analysis, then stable characters which distinguish stocks can be found using discriminant function procedures. Discriminant analysis is an objective means of identifying group membership and selecting characters; therefore, it may be preferable to principal component analysis for stock identification research.

RÉSUMÉ

Claytor, R.R. and H.R. MacCrimmon. 1986. Partitioning size from morphometric data: a comparison of five statistical procedures used in fisheries stock identification research. Can. Tech. Rep. Fish. Aquat. Sci. No. 1531: viii + 23 p.

Nous avons comparé cinq méthodes statistiques en usage courante pour standardiser des données morphométriques provenant de poissons de différentes tailles. Nos données comprenaient des classes sans chevauchement, avec chevauchement partiel et avec chevauchement complet de taille. Les méthodes utilisant la variation totale, plutôt que la variation intra-groupe, se sont le plus souvent avérées utiles pour standardiser les tailles, identifier les stocks et choisir les caractéristiques discriminatoires. Ces méthodes sont (i) l'analyse discriminante utilisant des mesures corrigées par le total des pentes combinées, et (ii) l'analyse

de composantes principales utilisant la matrice de covariance totale comme information de départ. Les autres procédures examinées, soit (iii) l'analyse discriminante utilisant les mesures corrigées par les pentes intra-groupes combinées, (iv) l'analyse de composantes principales utilisant la matrice de covariance intra-groupe ou (v) la méthode shear n'étaient pas utiles pour standardiser les données de taille de la série sans chevauchement. Il est conclu que les méthodes utilisant la variation totale sont à recommander si les stocks proposés sont formés de sous-groupes de différentes tailles.

Toutefois, s'il existe un chevauchement quelconque entre les tailles des stocks à comparer, les méthodes utilisant les analyses discriminantes avec mesures corrigées par les pentes totales combinées ou les pentes intra-groupes combinées, ou les méthodes utilisant l'analyse de composantes principales basées sur la matrice de covariance totale ou sur la méthode shear sont semblables. La correction tenant compte des tailles devra être effectuée avant l'analyse. Alors les caractéristiques servant à séparer les stocks pourront être identifiées en utilisant des méthodes d'analyse discriminante.

L'analyse discriminante est un moyen objectif pour identifier les groupes d'appartenance et pour choisir les caractéristiques discriminatoires. Donc elle est peut-être préférable à l'analyse de composantes principales en recherche sur l'identification de stocks.

INTRODUCTION

Morphometric characters are important for fish stock identification (Riddell and Leggett 1981; Meng and Stocker 1984; Winans 1984; MacCrimmon and Claytor 1985). However, change in shape resulting from increase in size must be considered in the analysis of morphometric data (Gould 1966). As fish growth is indeterminate, size can be considered a stock characteristic only when dealing with the same age groups among proposed stocks. Thus, size variation must often be partitioned so that decisions regarding stock structure can be made from characters independent of size (Thorpe 1983).

At present, two types of analyses and approaches are used to partition size in order to identify stocks from morphometric data. The two types of analyses are: (i) discriminant function analysis, which uses as input data characters adjusted for size by regression and (ii) principal component analysis, which uses a covariance matrix as input data. The two approaches used to obtain input data for either type of analysis are (i) utilization of the total variation in the data irrespective of group identity (total overall regression slopes and total covariance matrix) and (ii) utilization of the pooled within-group variation which requires group definition before partitioning size (pooled within-group slopes, pooled within-group covariance matrix and shear procedure).

As there is no consensus on which combination of analysis and approach is most appropriate for partitioning size, a variety of procedures has been employed in past studies. Five procedures currently used are: discriminant function analysis (Srivastava and Carter 1983) using characters adjusted for size by (i) the total overall regression slopes or (ii) the pooled within-group slopes derived using analysis of covariance (Snedecor and Cochran 1980) and principal component analysis (Morrison 1976) using (iii) the total covariance matrix, (iv) the pooled within-group covariance matrix or (v) the shear procedure (Humphries et al. 1981) to derive the principal components. As conclusions may depend on the procedure used, it is important to compare the effect each has on stock identification decisions.

This study is unique in comparing each of the above procedures on field data composed of subgroups likely to be encountered in a stock identification problem. The field data sets are organized as follows: (i) no overlap, (ii) partial overlap, and (iii) complete overlap in size between the groups comprising the stocks examined. This approach complements previous studies which have demonstrated the applicability of these procedures on single field data sets or have compared procedures using simulated data of predefined structure (Thorpe 1976; Humphries et al. 1981; Misra and Ni 1983; Thorpe 1983; Bookstein et al. 1985; Misra 1985). The results obtained in this study are important in deciding which procedures to use given a particular data structure and in assessing the similarity between studies which have employed different procedures.

MATERIALS AND METHODS

Juvenile Atlantic salmon were collected from the St. John River, New Brunswick, Canada, and the River Skealtar, North Uist, Scotland. Fish from these sites are recognized as distinct stocks based on morphometric, meristic and electrophoretic criteria (Claytor 1984).

The following eleven morphometric characters were measured from specimens preserved in formalin: head length, upper jaw length, distance between pectoral and pelvic fins, distance between pelvic and anal fins, gape width, head width, body width, head depth, body depth, and caudal peduncle depth, as defined by Trautman (1957), and standard length as defined by MacGregor and MacCrimmon (1977). These characters were found to be the most useful for separating Atlantic salmon stocks (Claytor 1984; MacCrimmon and Claytor 1985).

The characters were transformed to $\log_{10}$ values for all statistical analyses. The effectiveness of this transformation for removing mean/variance relationships and producing normally distributed variates was evaluated by visual inspection of residual versus predicted means, and variance versus mean scatterplots for each character by non-overlapping size group. Regression of variance versus mean values was used to assist visual inspections.

Residual versus predicted mean scatterplots were also used to identify outliers. Any value greater than three standard deviations from the mean was considered an outlier. All outliers were the result of data collection errors, and were removed prior to final tests for normality and all statistical analyses.

Three configurations of the data set were used in the statistical comparison described below. (1) Each stock was divided into seven non-overlapping size groups of 18-25 specimens each. A group was formed wherever a cluster of at least 20 (maximum 25) fish with similar standard lengths occurred in the data set. Exclusion of specimens with one or more missing values reduced the sample size of certain subgroups. (2) Each stock was divided into five partially overlapping size groups of seven specimens from each of three sequential groups from the non-overlapping data set. Sample sizes ranged from 19-21. (3) Size groups within each stock were combined to form completely overlapping data sets of 141-143 specimens for each stock (Fig. 1).

Five statistical procedures were compared using these data sets. The geographic stocks act as the control and the amount of overlap between subgroups acts as the treatments. In this way, it is simple to examine the effect of data structure on consistency of the statistical procedures in partitioning size, identifying stocks and character selection.

Two types of regression analysis were used to adjust the morphometric characters to a common standard length. The total overall regression slopes were derived irrespective of group origin and the separate and pooled within-group slopes were determined using analysis of covariance (ANACOVA). Thorpe (1976) provides a detailed

explanation of the differences between these slopes. Standard length was the independent (x) variable in the bivariate regression analysis (Draper and Smith 1981) and was the covariate in ANACOVA (Snedecor and Cochran 1980). Each of the other morphometric characters was the dependent (y) variable in 10 separate calculations. The SAS Proc GLM procedure (SAS 1979) was used to perform the regression and ANACOVA procedures. Homogeneity of the group slopes was tested using the F-test described by Sokal and Rohlf (1981, Chapter 14). Using the total overall and pooled within-group slopes, new variables, adjusted to the overall mean standard length, were obtained using the formula:

$$ACHR = OCHR - [SLOPE \times (SL - MSL)]$$

where ACHR equals the value of the size adjusted character, OCHR is the original value of the character, SLOPE is the slope of the character, SL is the standard length of each fish, and MSL is the overall mean standard length (Snedecor and Cochran 1980).

These adjusted characters were used as input to forward stepwise and forward followed by backward selection discriminant function analyses. This combination was used to test the character selection stability of the discriminant function (Dixon et al. 1981; Draper and Smith 1981; Srivastava and Carter 1983). The jackknife classification matrix was used to assess the accuracy of the discriminant functions in identifying geographic stocks. The jackknife procedure reduces the bias inherent in basing classification decisions on the data set which is being analyzed (Srivastava and Carter 1983). The kappa statistic was used to determine the improvement over chance of the classifications derived from discriminant function analysis (Titus et al. 1984).

Principal component scores derived from eigenvectors of the total covariance matrix, the pooled within-group covariance matrix, and the shear procedure have been calculated using the SAS procedure provided by Bookstein et al. (1985). In the shear procedure component scores derived from the total covariance matrix are used to identify groups. A second principal component analysis is then used to obtain the pooled within-group covariance matrix. The first eigenvector of this matrix is used as a within-group size vector (S). Two multiple regressions follow: (i) an optimal estimate of S ($\hat{S}$) is calculated by a regression of S against the eigenvectors of the total covariance matrix translated to zero mean and (ii) shape vectors independent of size are calculated by a regression of $\hat{S}$ and the non-centered total covariance matrix eigenvectors (Bookstein et al. 1985).

Size partitioning was tested by a regression of specimen scores for the proposed size-free characters against the size standard associated with each procedure. For discriminant function analysis, the first two canonical variable scores were regressed against standard length. The second and third components were regressed against the first of each principal component analysis. For example, the first component derived from the within-group covariance matrix was used as the size standard for the shear procedure. R-squared values were used as a relative measure of partitioning. Size was considered to be significantly partitioned if the regressions were not significant ($P > 0.001$).

Reduced character sets were formed using the most differentiated characters between stocks identified by principal component analyses using the total covariance matrix and the pooled within-group covariance matrix as input. The three principal component analyses were repeated using these reduced character sets. This was done to produce comparable results to discriminant analysis, which uses only the most differentiated characters to distinguish between groups. Only the completely overlapping data set was used in this portion of the analysis.

RESULTS

Log transformations successfully removed mean/variance relationships and produced normally distributed variates. There was no mean/variance relationship apparent upon visual inspection of scatterplots. Variance versus mean regression analyses confirmed non-significant relationships as all P-values were > 0.850. A horizontal band of residuals versus predicted means supported the conclusion drawn from mean/variance relationships. The assumptions necessary to optimize discriminant function analysis were satisfied.

Discriminant function analysis using characters adjusted for size by the total overall regression slopes was consistent in partitioning size, identifying stocks, and selecting characters from all data sets. Only this procedure significantly partitioned size from all data sets (Table 2). Size was not a factor in sorting the subgroups on the first or second canonical variable (Fig. 2). The non-overlapping and partially overlapping subgroups were classified correctly by size 30% of the time. This was 16-33% greater than would have occurred by chance (Table 3). Geographic separation of the subgroups occurred on the first canonical variable, accounted for 85% of the variance, and was greater than 98% accurate in identifying stock origin. Body depth, head depth and body width were the characters identified as important in distinguishing between the geographic stocks (Tables 3, 4; Fig. 2).

Discriminant function analysis using characters adjusted for size by the pooled within-group slopes significantly partitioned size from the partially and completely overlapping, but not from the non-overlapping data sets. The non-overlapping subgroups were ordered by size regardless of geographic origin on the first and second canonical variables (Table 2; Fig. 3). Identification of the non-overlapping subgroups by size exceeded 56%, which was 47-60% better than chance (Table 3). Hence, the geographic separation observed between the non-overlapping subgroups was not independent of size. The characters identified as important for differentiating stocks using the pooled within-group slopes were equivalent to those using the total overall slopes, but only for the partially and completely overlapping data sets (Fig. 3). For the non-overlapping data set, the pooled within-group slope (0.57) of head depth versus standard length, one of the important

characters distinguishing between stocks, was not parallel to the total overall regression slope (0.91). The pooled within-group slopes derived using ANACOVA on the partially (0.88) and completely (0.89) overlapping data sets were, however, parallel to the total overall slope. The R-squared values of these regressions were > 0.97.

Character correlations were much lower after than before size adjustments using total overall regression slopes. Only three exceptions occurred, each in the non-overlapping data set (gape width, pelvic to anal fin distance; gape width, pectoral to pelvic fin distance; upper jaw length, pelvic to anal fin distance). These characters were weakly correlated, even though they were all highly correlated to standard length prior to size adjustment (Table 5). Correlations were also reduced when ANACOVA was used to correct for size.

The first component of each principal component analysis was interpreted as a size factor regardless of data set for two reasons. First, all signs for the component coefficients were positive and second, loadings were identical among traits using the total covariance matrix and nearly so using the within-group covariance matrix to derive components (Table 6).

Principal component analysis using the total covariance matrix as input was consistent in partitioning size from all data sets. The second and third components were interpreted as size-free for all data sets, as R-squared values were low and no size relationship was apparent when these components were plotted against the first or size component (Table 2; Fig. 4, 5). However, size was significantly partitioned only from the partially overlapping data set (Table 2). The identification accuracy of the geographic stocks was less using this principal component analysis than the discriminant function procedures for the non- and completely, but not the partially overlapping data set (compare Figs. 2 and 3 to 4 and 5). Body depth, head depth, body width, and distance between the pelvic and anal fins were consistently identified as important in distinguishing between the geographic stocks (Table 7; Figs. 4, 5).

Principal component analysis using the pooled within-group covariance matrix significantly partitioned size only from the completely overlapping data set (Table 2). The second component retained elements of size when the non-overlapping data set was analyzed and the third component retained elements of size with the partially overlapping data set (Table 2; Fig. 6). The identification accuracy of the geographic stocks was lower using this analysis than principal component analysis using the total covariance matrix (compare Figs. 4, 5, 6). No consistent character set was identified as important using principal component analysis based on the pooled within-group covariance matrix (Table 7; Fig. 6).

Principal component analysis using the shear procedure did not partition size from the non-overlapping data set. The second and third component scores were highly correlated with size (Table 2; Fig. 7). The complete separation of stocks in the non-overlapping data set is size dependent and therefore not meaningful (Fig. 7). However, when the partially and completely overlapping data sets were analyzed, the second and third components were almost identical to the loadings determined using the total covariance matrix. Hence, for these data sets identical characters were important using the shear and total covariance matrix procedures (Table 7; Figs. 4, 5, 7).

Total separation of stocks was achieved using the total covariance matrix and shear procedure to derive principal components using only body depth, body width, head depth, and distance between the pelvic and anal fins in the completely overlapping data set. Partial separation of stocks occurred using the pooled within-group covariance matrix to derive components from this reduced character set (Fig. 8). These characters were identified as important for distinguishing between stocks using the total covariance matrix to derive principal components (Figs. 4, 5). Separation of stocks was incomplete using gape width, distance between the pectoral and pelvic fins, and distance between the pelvic and anal fins to analyze the completely overlapping data set (Fig. 9). These characters were identified as most differentiated between stocks in the completely overlapping data set using the pooled within-group covariance matrix to derive the principal components (Fig. 6). As before, principal components derived using the total covariance matrix and shear procedure were nearly identical (Table 8).

DISCUSSION

The procedures which utilized total rather than within-group variation were most consistent in partitioning size. This result contrasts with authors preferring or recommending procedures requiring designation of groups before adjusting for size. For example, Misra and Ni (1983) and Misra (1985) recommend using discriminant analysis and the pooled within-group slopes, Thorpe (1976, 1983) prefers principal component analysis using the pooled within-group covariance matrix; and Humphries et al. (1981) and Bookstein et al. (1985) are proponents of the shear procedure.

Extrapolation error may occur when ANACOVA is used to calculate pooled within-group slopes if there is little or no overlap among sample groups with respect to the covariate. In this situation, inferences regarding relationships beyond the data points are suspect. Snedecor and Cochran (1980) identify extrapolation error as one of the potential problems associated with using ANACOVA to adjust non-overlapping data sets.

The contrast between procedures using total rather than within-group variation partly results from this extrapolation error. Thorpe (1976) recommends that the pooled within-group slope be used exclusively when employing regression procedures to partition size from morphometric data because the pooled within-group slope may be different from the total slope. However, this study shows that the difference in slopes occurs as a result of the data structure and not because of an inherent bias in using the total slope.

Exceptions to the use of the total slope will depend on the structure of the data and can easily be tested by visual inspection, F-tests, and R-squared comparisons (Draper and Smith 1981; Sokal and Rohlf 1981). There are certain cases where

ANACOVA and using the pooled or separate within-group slopes will be more appropriate than using the total slope. For example, one or more groups may have a significantly different slope from the total or pooled within-group slope or the separate within-group slopes may be parallel but widely separated (Bookstein et al. 1985).

The prevalence of these exceptions appears to be low in stock investigations. The pooled within-group slope is commonly used when ANACOVA has been employed to adjust for size (Riddell and Leggett 1981; Ihssen et al. 1981; Beacham 1984; MacCrimmon and Claytor 1985). Separate within-group slopes are not likely to be widely separated in intraspecies stock investigations. Cases where size correction by any means may be inappropriate (i.e. there is no correlation between morphometric characters and standard length) are also expected to be rare in stock investigations.

Analogous arguments pertaining to extrapolation error and the contrast between total and within-group variation also apply to the multivariate case. For example, the high gape width and low pelvic to anal fin first component scores of the within-group principal component analysis probably result from the unusually low correlations between these characters in the non-overlapping data set (Table 5A). These characters were highly correlated to standard length in all and to each other in the remaining data sets (Table 5). As with regression, this sensitivity of within-group procedures to unusual relationships among characters may explain the failure of principal component analysis using the pooled within-group covariance matrix and shear procedure to partition size from the non-overlapping data set.

When adjustments for size are made prior to analysis, stable and interpretable characters distinguishing between stocks are identified using discriminant analysis. This result contrasts with Humphries et al. (1981), who found characters selected using discriminant analysis to be inconsistent. All discriminant function and principal component analyses which successfully partitioned size identified the same characters as differentiating between stocks (compare Figs. 2-5). Hence, the relationships identified by these procedures are equally interpretable. The stability of discriminant analysis is supported by the finding that the forward stepwise and the forward followed by the backward procedures identified the same character sets as most differentiated between stocks. When a data set is composed of highly correlated characters, the order in which the variables are entered into the model may influence final character selection (Draper and Smith 1981). Adjustment for size using regression or principal component analysis reduces correlations between characters (Table 3), enabling stable character sets to be identified. The suggestion that discriminant analysis does not produce stable and interpretable character sets seems not to apply to adjusted characters.

Using only the most differentiated characters improved the stock identification of principal component analysis. However, a disadvantage of this analysis is that the choice of stock division and character selection is subjective. In contrast, discriminant analysis is more objective for identifying group membership and selecting important characters.

Finally, it has been suggested that partitioning size by excluding the first and designating the remaining principal components as size-free is an arbitrary decision (Humphries et al. 1981 and references therein). The results from the three data sets examined support this suggestion. None of the principal component procedures produced significantly size-free vectors in both the second and third components for all analyses (Table 2). Therefore, discriminant analysis using overall regression slopes to partition size may be preferable for stock identification problems.

CONCLUSIONS

If there is at least some size overlap among the groups examined, past investigations which utilize discriminant function analysis with total or pooled within-group slopes are probably similar to those employing the more recently developed principal component analysis with the total covariance matrix and shear procedure. However, if there is no size overlap among the groups examined, then discriminant function analysis using the pooled within-group slopes and principal component analysis using the shear procedure do not partition size. This result emphasizes the importance of reporting the size structure of the data set being examined.

Investigations using principal component analysis based on the within-group covariance matrix may be different from the other four procedures. This procedure produced appreciably different results than the other four with respect to size partitioning, accuracy of stock identification, and character selection.

The contrasts between this investigation and others (Thorpe 1976; Humphries et al. 1981; Misra and Ni 1983; Thorpe 1983; Bookstein et al. 1985; Misra 1985) emphasize that size partitioning should not be assumed without testing. Regressing the canonical variables or principal components against an appropriate size standard is an easy method of testing this assumption. Such examinations will be useful for developing more robust and powerful procedures for partitioning size from morphometric data.

ACKNOWLEDGEMENTS

M. Chadwick, M. Blouw, P. Ihssen, G. Winans, and R. J. Hines provided many helpful suggestions in review of early drafts. We wish to thank J. McAskill and N. Campbell for collecting specimens. Financial support was provided by Fisheries and Oceans subvention and NSERC operating grants to H. R. MacCrimmon. We would also like to thank the referees for many useful suggestions.

REFERENCES

Beacham, T. D. 1984. Age and morphology of chum salmon in southern British Columbia. Trans. Am. Fish. Soc. 113:717-736.

Bookstein, F. L., B. Chernoff, R. L. Elder, J. M. Humphries, G. R. Smith, and R. E. Strauss. 1985. Morphometrics in evolutionary biology: the geometry of size and shape change with examples from fishes. Academy of Natural Sciences of Philadelphia, Philadelphia, Pa. 277 p.

Claytor, R. R. 1984. The taxonomy of Atlantic salmon, Salmo salar Linnaeus: a species description relative to the stock concept. M.Sc. thesis, University of Guelph, Guelph, Ontario. 136 p.

Dixon, W. J., M. B. Brown, L. Engelman, J. W. Frane, M. A. Hill, R.I. Jennrich, and J. D. Toporek. 1981. BMDP statistical software. University of California Press, Berkeley, Ca. 725 p.

Draper, N. R., and H. Smith. 1981. Applied regression analysis. John Wiley and Sons, Inc., New York. 709 p.

Gould, S. J. 1966. Allometry and size in ontogeny and phylogeny. Biol Rev. 41:587-640.

Humphries, J. M., F. L. Bookstein, B. Chernoff, G. R. Smith, R. L. Elder, and S. G. Poss. 1981. Multivariate discrimination by shape in relation to size. Syst. Zool. 30:291-308.

Ihssen, P. E., D. O. Evans, W. J. Christie, J. A. Reckahn, and R.L. DesJardine. 1981. Life history, morphology, and electrophoretic characteristics of five allopatric stocks of lake whitefish (Coregonis clupeaformis) in the Great Lakes region. Can. J. Fish. Aquat. Sci. 38:1790-1807.

MacCrimmon, H. R., and R. R. Claytor. 1985. Meristic and morphometric identity of Baltic stocks of Atlantic salmon (Salmo salar). Can. J. Zool. 63:2032-2037.

MacGregor, R. B., and H. R. MacCrimmon. 1977. Evidence of genetic and environmental influences on meristic variation in the rainbow trout, Salmo gairdneri Richardson. Environ. Biol. Fish. 2:25-33.

Meng, H. J., and M. Stocker. 1984. An evaluation of morphometrics and meristics for stock separation of Pacific herring (Culpea harengus pallasi). Can. J. Fish. Aquat. Sci. 41:414-422.

Misra, R. K., and I-H. Ni. 1983. Distinguishing beaked redfishes (deepwater redfish, Sebastes mentella and Labrador redfish, S. fasciatus) by discriminant analysis (with covariance) and multivariate analysis of covariance. Can. J. Fish. Aquat. Sci. 40:1507-1511.

Misra, R. K. 1985. Quadratic discriminant analysis with covariance for stock delineation and population differentiation: a study of beaked redfishes (Sebastes mentella and S. faciatus). Can. J. Fish. Aquat. Sci. 42:1672-1676.

Morrison, D. F.. 1976. Multivariate statistical methods. McGraw Hill, Inc., New York. 415 p.

Riddell, B. E., and W. C. Leggett. 1981. Evidence of an adaptive basis for geographic variation in body morphology and time of downstream migration of juvenile Atlantic salmon (Salmo salar). Can. J. Fish. Aquat. Sci. 38:308-320.

SAS. 1979. SAS user's guide 1979 edition. SAS Institute Inc. Box 8000, Cary, North Carolina. 494 p.

Snedecor, G. W., and W. G. Cochran. 1980. Statistical methods. The Iowa State University Press, Ames. IA. 593 p.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. 2nd ed., W. H. Freeman and Co., San Francisco, Ca. 859 p.

Srivastava, M. S., and E. M. Carter. 1983. An introduction to applied multivariate statistics. North-Holland Publishing Co., New York. 394 p.

Thorpe, R. S. 1976. Biometric analysis of geographic variation and racial affinities. Biol. Rev. 51:407-452.

Thorpe, R. S. 1983. A review of the numerical methods for recognizing and analyzing racial variation. In J. Felsenstein (ed.), Numerical Taxonomy: 404-423. Berlin: Springer-Verlag.

Titus, U., J. A. Mosher, and B. K. Williams. 1984. Chance-corrected classification for use in discriminant analysis: ecological applications. Am. Mid. Nat. 111:1-7.

Trautman, M. B. 1957. The fishes of Ohio. Ohio State University Press, Columbus, OH. 683 p.

Winans, G. A. 1984. Multivariate morphometric variability in Pacific salmon: technical demonstration. Can. J. Fish. Aquat. Sci. 41:1150-1159.

Table 1. Acronyms used for morphometric characters, geographic stocks, data sets, analyses, and symbols in subsequent tables and figures.

Morphometric characters	
SL	Standard length
HL	Head length
UJL	Upper jaw length
POPB	Pectoral to pelvic fin distance
PBAB	Pelvic to anal fin distance
GW	Gape width
HW	Head width
BW	Body width
HD	Head depth
BD	Body depth
CPD	Caudal peduncle depth

Data sets and analyses	
NO	Non-overlapping
PO	Partially overlapping
CO	Completely overlapping
DFA	Discriminant function analysis
PCA	Principal component analysis
TOTAL	Total variation procedures
WITHIN	Within variation procedures
C	Canada
S	Scotland

Symbols	
—— HL	Character vectors represent addition of the coefficients of the most differentiated characters for the canonical variables or principal components shown.
☐	Borders within each figure enclose all specimens for each stock.
1s, 1c	Numbers refer to mean canonical or principal component score of each subgroup, subscript refers to geographic stock.

Table 2. R-squared values for regressions of size-free characters against size standard associated with each procedure. * indicates a non-significant regression (P > 0.001). Variable refers to either canonical variable or principal component. Acronyms as listed in Table 1.

Procedure	Variable	Data set		
		Non-overlap	Partial	Complete
Total DFA	1	0.0*	0.0*	0.0*
	2	0.0*	0.0*	N/A
Total PCA	2	7.4	0.1*	7.4
	3	17.7	0.2*	17.7
Within DFA	1	59.3	1.2*	1.1*
	2	24.9	11.9	N/A
Within PCA	2	88.7	0.1*	4.1*
	3	12.5	39.9	0.7*
SHEAR	2	80.6	0.0*	6.7
	3	89.5	36.6	0.2*

Table 3. Kappa and average observed percentage of specimens correctly identified to size groups and geographic stocks by discriminant function analysis. Values for complete data set are for geographic stocks. Acronyms as listed in Table 1.

Data set	Observed percentage size groups	Kappa	95% C.I. kappa	Geographic stocks
Non-overlapping				
Total	30.2	24.8	16.5- 33.1	98.2
Within	56.8	53.5	47.2- 59.8	98.2
Partial				
Total	31.9	24.3	17.3- 31.3	98.6
Within	35.7	28.6	21.6- 35.6	99.5
Complete				
Total	N/A	98.6	96.6-100.6	99.3
Within	N/A	99.3	97.9-100.7	99.6

Table 4. The standardized coefficients of the first two canonical variables from the discriminant analysis using characters adjusted for size by total and pooled within-group regression slopes. Blank indicates that the character did not enter the model. Acronyms as listed in Table 1.

A) First canonical variable

	Total			Within		
Character	NO	PO	CO	NO	PO	CO
BD	- 0.5	- 0.5	- 0.5	- 0.2	- 0.4	- 0.5
PBAB			- 0.1		- 0.1	- 0.1
POPB		- 0.1			- 0.1	
HW				- 0.2		
GW				0.0		
HL	- 0.1		- 0.2	0.3		- 0.2
BW	0.4	0.4	0.4	0.1	0.4	0.4
HD	0.6	0.5	0.6	0.7	0.5	0.6
Percent variance	85	93	100	71	89	100

B) Second canonical variable

	Total		Within	
Character	NO	PO	NO	PO
POPB		- 0.7		- 1.1
BD	- 0.7	- 0.6	- 0.6	- 0.4
HD	0.0	- 0.3	0.1	- 0.3
HW			0.1	
GW	-		0.2	
PBAB				0.5
BW	- 0.3	0.1	0.7	0.4
HL	0.9		- 0.9	
Percent variance	11	4	24	6

Table 5. Partial correlation coefficients between morphometric characters before and after adjusting for size using total regression slopes for the non-overlapping, partial overlapping and complete overlapping data sets. Upper triangular matrix after size adjustment, lower triangular prior to size adjustment. Acronyms as listed in Table 1.

A) Non-overlap

	Character										
Character	SL	HL	UJL	POPB	PBAB	GW	HW	BW	HD	BD	CPD
SL	X	-0.16	-0.10	0.11	0.02	-0.01	0.02	0.01	-0.24	-0.04	-0.07
HL	0.58	X	0.28	-0.09	-0.09	0.16	0.21	0.19	0.29	0.17	0.22
UJL	0.46	0.46	X	-0.07	-0.17	0.25	0.16	0.16	0.24	0.21	0.17
POPB	0.66	0.34	0.26	X	0.13	-0.26	0.29	0.06	-0.10	0.01	-0.01
PBAB	0.52	0.24	0.11	0.43	X	-0.24	0.01	-0.08	-0.10	-0.09	-0.10
GW	0.31	0.31	0.35	0.02	-0.04	X	0.27	0.41	0.28	0.37	0.28
HW	0.53	0.46	0.36	0.53	0.28	0.38	X	0.58	0.26	0.44	0.27
BW	0.50	0.43	0.36	0.37	0.20	0.49	0.69	X	0.25	0.43	0.19
HD	0.36	0.41	0.36	0.19	0.11	0.37	0.41	0.39	X	0.39	0.22
BD	0.51	0.41	0.39	0.35	0.20	0.46	0.59	0.58	0.50	X	0.45
CPD	0.54	0.46	0.37	0.36	0.21	0.40	0.48	0.41	0.37	0.60	X

B) Partial overlap

	Character										
Character	SL	HL	UJL	POPB	PBAB	GW	HW	BW	HD	BD	CPD
SL	X	0.04	-0.01	-0.11	0.15	-0.07	-0.14	-0.06	-0.16	-0.02	-0.11
HL	0.97	X	0.40	-0.09	-0.18	0.10	0.23	0.12	0.34	0.15	0.22
UJL	0.93	0.94	X	-0.01	-0.23	0.28	0.23	0.15	0.33	0.21	0.21
POPB	0.96	0.92	0.89	X	0.29	-0.20	0.35	0.13	-0.06	0.02	0.01
PBAB	0.95	0.91	0.86	0.94	X	-0.26	0.04	-0.08	-0.21	-0.03	-0.13
GW	0.77	0.77	0.78	0.70	0.69	X	0.35	0.47	0.33	0.46	0.38
HW	0.91	0.91	0.88	0.91	0.88	0.79	X	0.61	0.32	0.48	0.34
BW	0.92	0.90	0.87	0.89	0.86	0.83	0.94	X	0.30	0.55	0.28
HD	0.93	0.94	0.91	0.88	0.87	0.80	0.90	0.90	X	0.43	0.30
BD	0.92	0.91	0.89	0.89	0.88	0.83	0.92	0.93	0.92	X	0.51
CPD	0.95	0.93	0.91	0.90	0.89	0.81	0.91	0.90	0.92	0.94	X

C) Complete overlap

	Character									
Character	SL	HL	UJL	POPB	PBAB	GW	HW	BW	HD	BD	CPD
SL	X	-0.03	-0.05	0.05	0.09	-0.08	-0.03	-0.09	-0.18	0.03	-0.02
HL	0.99	X	0.39	-0.09	-0.15	0.06	0.17	0.05	0.25	0.01	0.14
UJL	0.98	0.98	X	-0.02	-0.20	0.23	0.21	0.10	0.28	0.16	0.15
POPB	0.99	0.98	0.97	X	0.17	-0.12	0.36	0.11	-0.01	0.16	0.05
PBAB	0.98	0.97	0.96	0.98	X	-0.17	0.03	-0.03	-0.11	-0.01	-0.04
GW	0.94	0.94	0.94	0.93	0.92	X	0.41	0.52	0.36	0.53	0.38
HW	0.98	0.98	0.97	0.98	0.97	0.95	X	0.59	0.34	0.53	0.34
BW	0.98	0.97	0.97	0.97	0.96	0.96	0.98	X	0.30	0.54	0.29
HD	0.99	0.98	0.98	0.98	0.97	0.95	0.98	0.98	X	0.43	0.25
BD	0.98	0.97	0.97	0.98	0.97	0.96	0.98	0.98	0.98	X	0.50
CPD	0.99	0.98	0.98	0.98	0.97	0.95	0.98	0.98	0.98	0.98	X

Table 6. First principal component coefficients derived using the total covariance matrix and the within-group covariance matrix from all three data sets. Acronyms as listed in Table 1.

Character	Total covariance			Within covariance		
	NO, PO, CO			NO	PO, CO	
SL	0.3			0.2	0.3	
HL	0.3			0.2	0.3	
UJL	0.3			0.2	0.3	
POP3	0.3			0.2	0.3	
PBAB	0.3			0.1	0.3	
GW	0.3			0.6	0.3	
HW	0.3			0.3	0.3	
BW	0.3			0.4	0.3	
HD	0.3			0.2	0.3	
BD	0.3			0.4	0.3	
CPD	0.3			0.3	0.3	
Percent variance	96.0	94.2	96.0	44.8	88.9	97.4

Table 7. The second and third principal component coefficients derived using the total covariance matrix, within-group covariance matrix and shear procedure for A) Non-overlapping data set; B) Partial overlapping data set; and C) Complete overlapping data set. Acronyms as listed in Table 1.

A) Non-overlap

	Principal component					
	2			3		
Character	Total	Shear	Within	Character	Total	Shear	Within
GW	-0.6	-0.4	-0.6	GW	-0.5	-0.4	-0.4
HD	-0.3	-0.2	0.0	BD	-0.5	-0.3	0.2
BW	-0.3	-0.2	0.1	PBAB	-0.2	-0.1	-0.8
HW	0.0	0.1	0.2	CPD	-0.1	0.0	0.1
UJL	0.0	0.1	0 0	HW	0.0	0.0	0.2
CPD	0.1	0.2	0.1	POPB	0.0	0.1	0.1
HL	0.1	0.2	0.1	SL	0.1	0.2	-0.1
BD	0.1	0.2	0.1	BW	0.1	0.2	0.2
SL	0.2	0.2	0.2	HL	0.3	0.3	0.1
POPB	0.3	0.4	0.4	UJL	0.3	0.3	0.2
PBAB	0.5	0.5	0.5	HD	0.5	0.5	0.2
Percent variance	2.2	N/A	19.0		0.5	N/A	8.1

B) Partial overlap

	Principal component 2				Principal component 3		
Character	Total	Shear	Within	Character	Total	Shear	Within
GW	-0.7	-0.7	-0.8	GW	-0.5	-0.5	0.1
HD	-0.3	-0.3	0.0	BD	-0.4	-0.4	0.2
BW	-0.3	-0.3	-0.1	PBAB	-0.2	-0.2	0.3
HW	0.0	0.0	0.0	CPD	-0.1	-0.1	-0.1
UJL	0.0	0.0	0.0	HW	0.0	0.0	0.3
CPD	0.1	0.1	0.0	POPB	0.0	0.1	0.2
HL	0.1	0.1	0.1	SL	0.1	0.1	-0.1
BD	0.1	0.1	-0.1	BW	0.2	0.2	0.4
SL	0.2	0.2	0.2	HL	0.2	0.2	-0.3
POPB	0.3	0.3	0.3	UJL	0.3	0.3	-0.6
PBAB	0.5	0.5	0.4	HD	0.5	0.6	-0.3
Percent variance	3.5	N/A	4.7		0.6	N/A	1.5

C) Complete overlap

	Principal component 2				Principal component 3		
Character	Total	Shear	Within	Character	Total	Shear	Within
GW	-0.6	-0.6	-0.8	GW	-0.5	-0.5	0.1
HD	-0.3	-0.3	0.0	BD	-0.5	-0.4	0.2
BW	-0.3	-0.3	-0.1	PBAB	-0.2	-0.2	0.5
HW	0.0	0.0	0.0	CPD	-0.1	-0.1	0.0
UJL	0.0	0.0	0.1	HW	0.0	0.0	0.1
CPD	0.1	0.1	0.0	POPB	0.0	0.0	0.1
HL	0.1	0.1	0.1	SL	0.1	0.1	-0.1
BD	0.1	0.1	-0.1	BW	0.1	0.1	0.3
SL	0.2	0.2	0.2	HL	0.3	0.3	-0.4
POPB	0.3	0.3	0.3	UJL	0.3	0.3	-0.7
PBAB	0.5	0.5	0.4	HD	0.5	0.5	-0.2
Percent variance	2.2	N/A	1.0		0.5	N/A	0.4

Table 8. Coefficients based on characters identified as most differentiated between stocks by principal component analysis of the entire character set. A) The four most differentiated characters identified by principal component analysis using the total covariance matrix. B) The three most differentiated characters identified by principal component analysis using the within-group covariance matrix. Acronyms as listed in Table 1.

A) Four most differentiated characters identified by the total covariance matrix

	Principal component					
	2				3	
Character	Total	Shear	Within	Character	Total	Shear	Within
HD	-0.5	-0.6	-0.2	PBAB	-0.5	-0.5	-0.1
BW	-0.5	-0.5	-0.3	HD	-0.3	-0.3	0.7
BD	0.2	0.2	-0.4	BW	0.1	0.0	-0.7
PBAB	0.7	0.7	0.8	BD	0.8	0.8	0.1

B) Three most differentiated characters identified by the within-group covariance matrix

	Principal component					
	2				3	
Character	Total	Shear	Within	Character	Total	Shear	Within
GW	-0.8	-0.8	-0.8	PBAB	-0.6	-0.6	-0.7
POPB	0.3	0.3	0.3	GW	-0.1	-0.1	-0.1
PBAB	0.5	0.5	0.5	POPB	0.8	0.8	0.7

Fig. 1. The organization of the Canadian and Scottish stocks into data sets composed of non-overlapping, partially and completely overlapping size subgroups.

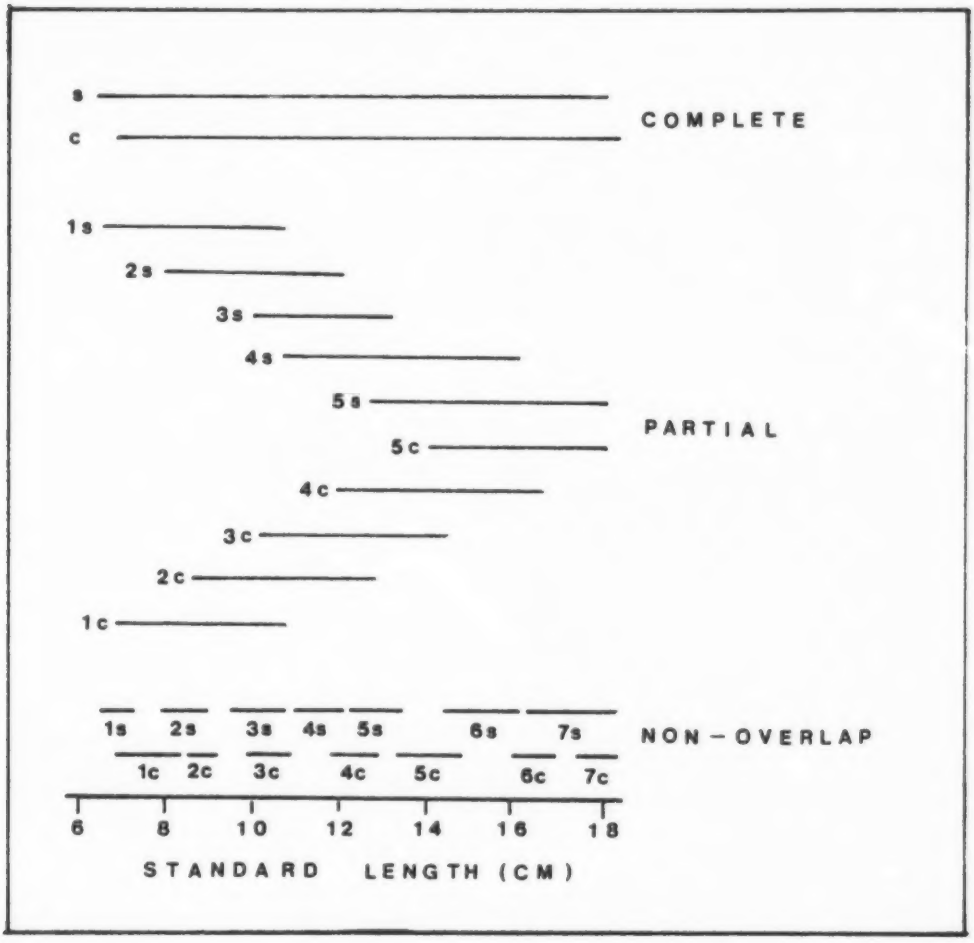

Fig. 2 Discriminant analysis using characters adjusted for size by the total overall regression slopes for all three data sets. Acronyms as listed in Table 1.

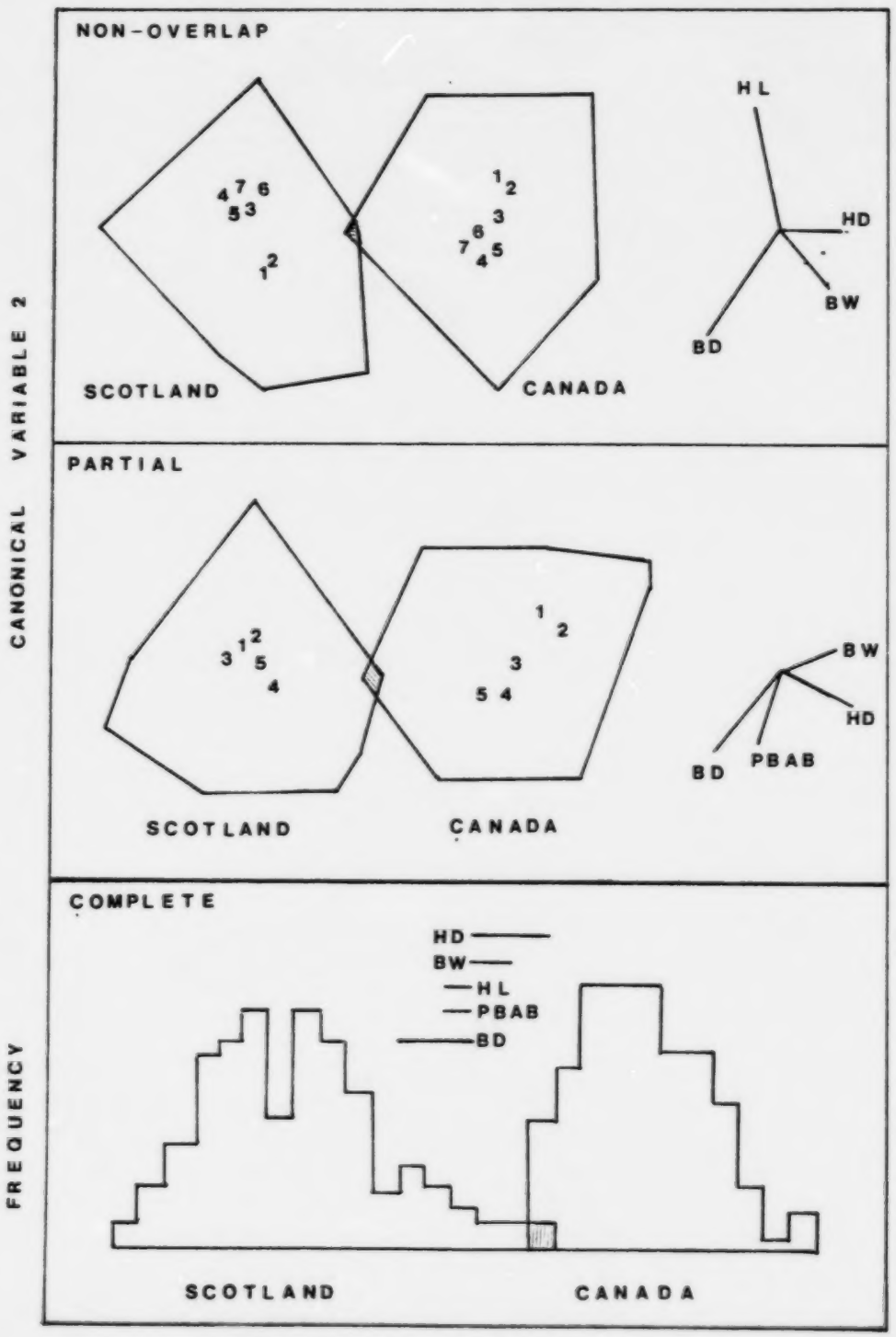

Fig. 3 Discriminant analysis using characters adjusted for size by the pooled within-group slopes for all three data sets. Acronyms as listed in Table 1.

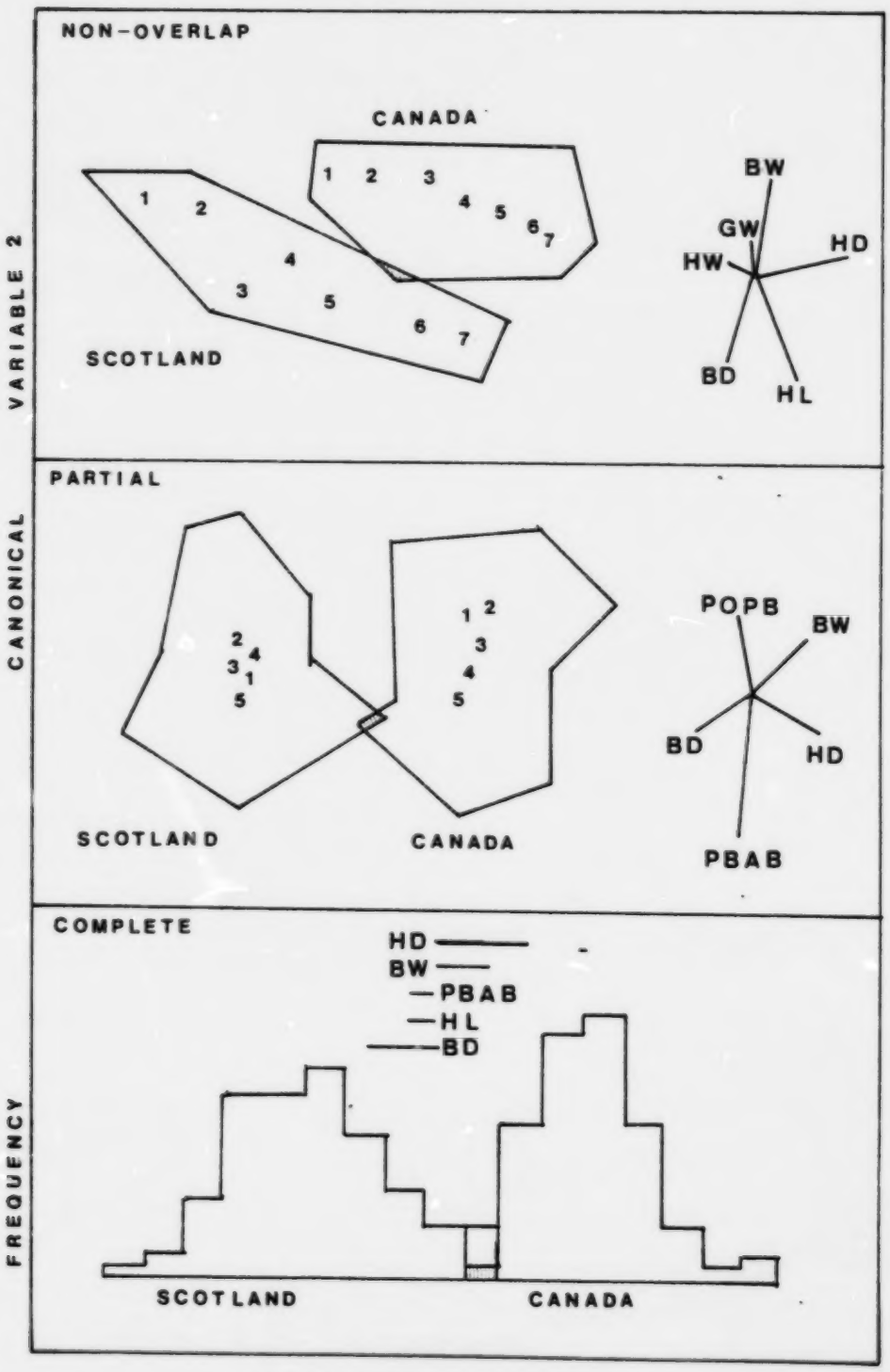

Fig. 4 Principal component analysis using the total covariance matrix for non-overlapping and completely overlapping data sets. Acronyms as listed in Table 1.

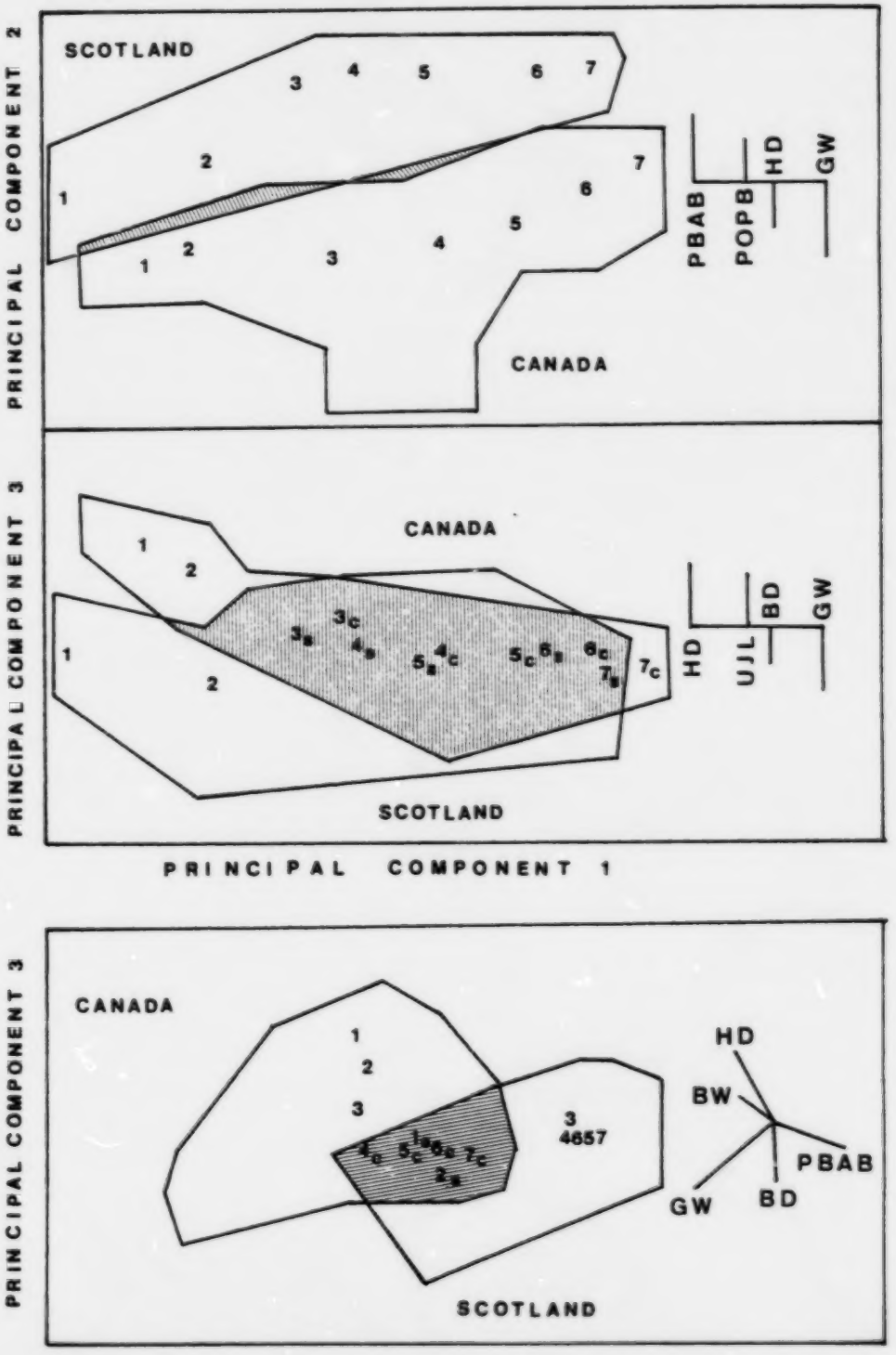

Fig. 5 Principal component analysis using the total covariance matrix for partial overlapping data set. Acronyms as listed in Table 1.

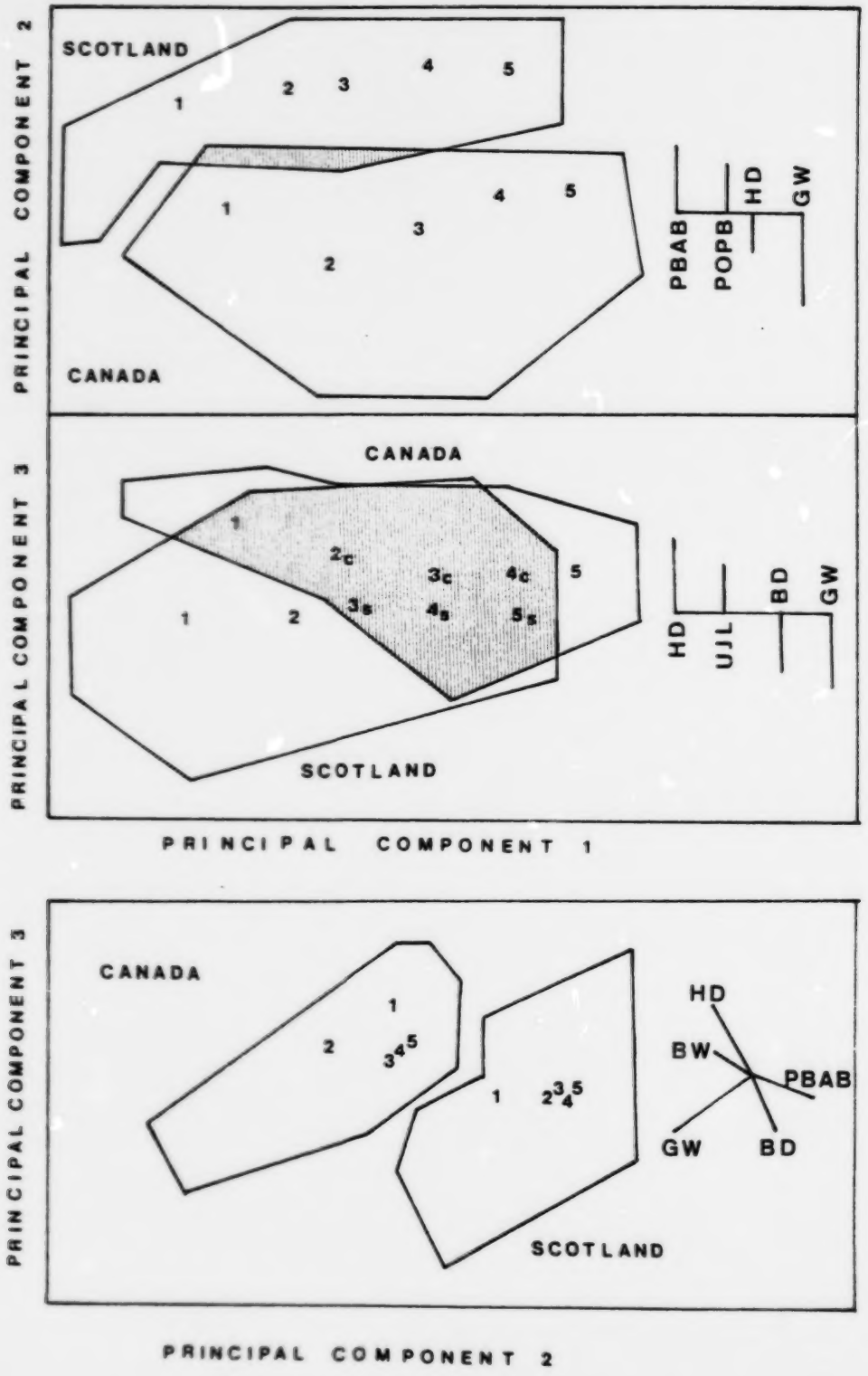

Fig. 6 Principal component analysis using the pooled within-group covariance matrix for all three data sets. Acronyms as listed in Table 1.

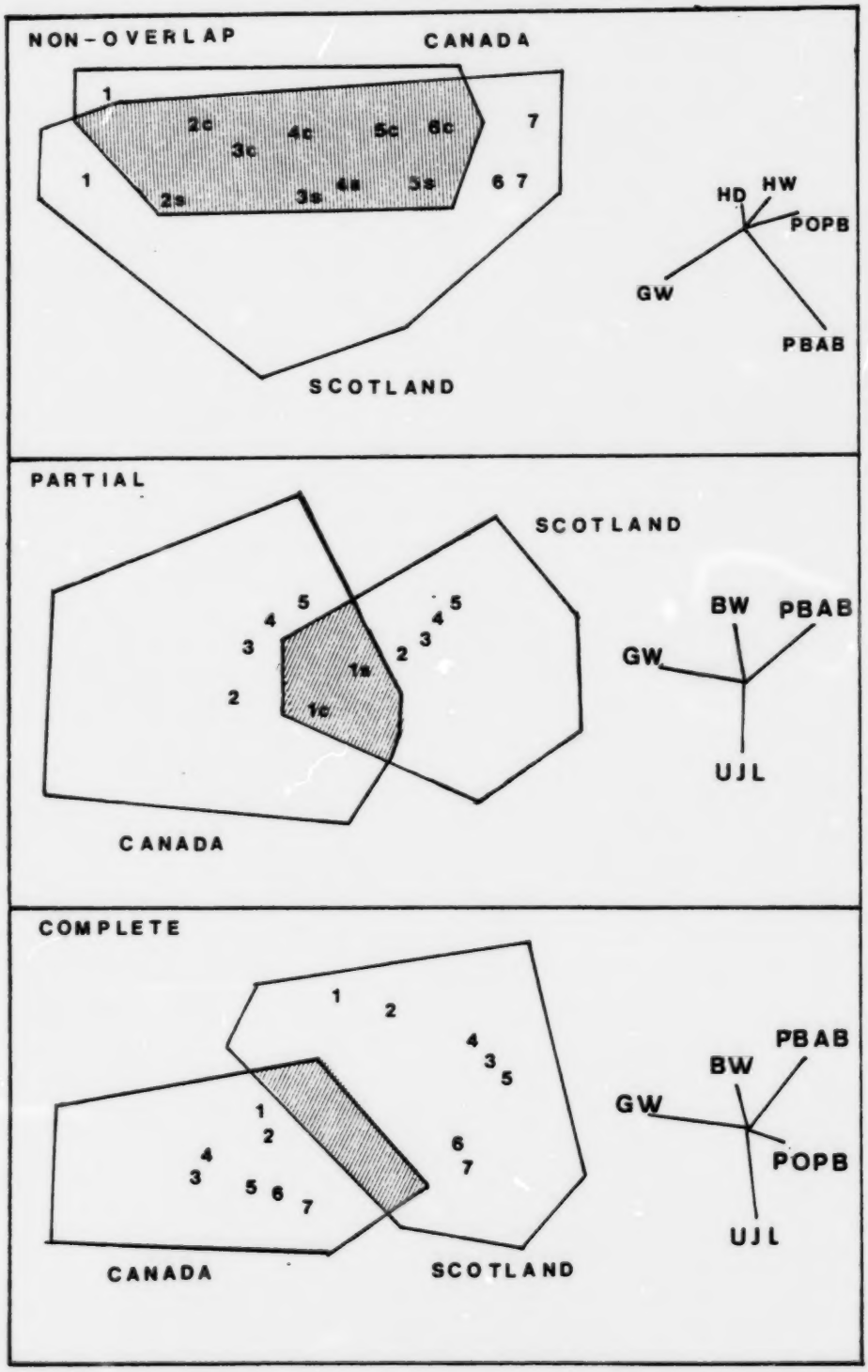

Fig. 7 Principal component analysis using the shear procedure for non-overlapping data set. Acronyms as listed in Table 1.

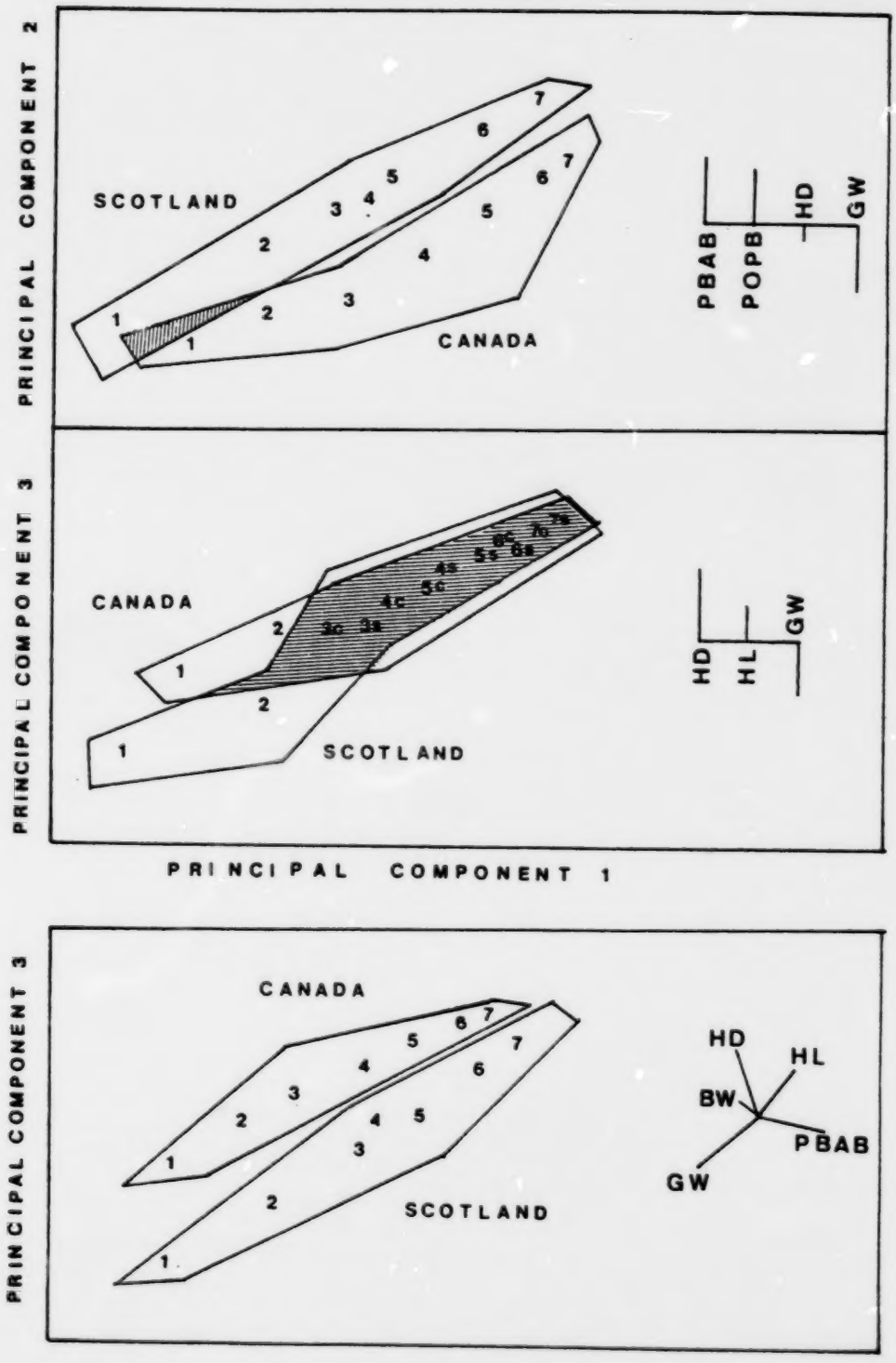

Fig. 8 Principal component analysis of the completely overlapping data set using the four most differentiated characters identified by principal component analysis using the total covariance matrix. Acronyms as listed in Table 1.

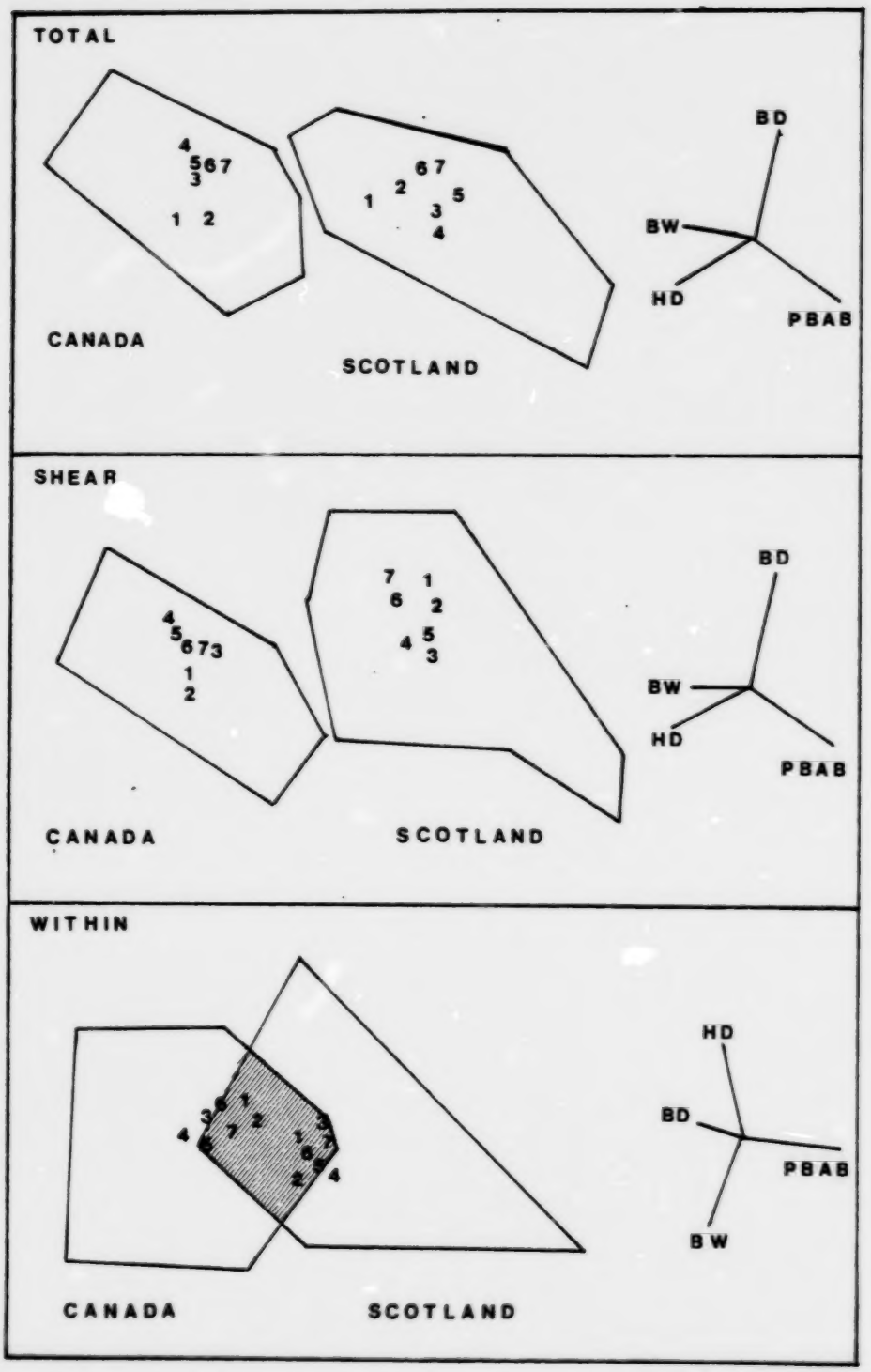

Fig. 9 Principal component analysis of the completely overlapping data set using the three most differentiated characters identified by principal component analysis using the within covariance matrix. Acronyms as listed in Table 1.